

CLAIMS

I claim:

1. A method of dynamically modifying resources within a compute environment, the method comprising:
 - receiving a request for resources in the compute environment;
 - monitoring events after receiving the request for resources; and
 - based on the monitored events, dynamically modifying at least one of the request for resources and the compute environment.
2. The method of claim 1, wherein the compute environment is one of a compute farm, a cluster environment and a grid environment
3. The method of claim 1, wherein the request for resources is a request for consumption resources.
4. The method of claim 1, wherein the request for resources is a request for provisioning services.
5. The method of claim 1, wherein the request for resources is a request to process a batch job.
6. The method of claim 1, wherein the request for resources is a request for direct volume access.
7. The method of claim 1, wherein the request for resources is a request for a virtual private cluster.
8. The method of claim 1, wherein monitoring events after receiving the request for resources further comprises monitoring the compute environment.
9. The method of claim 1, wherein monitoring events after receiving the request for resources further comprises monitoring to determine if a party submitting the request has submitted a job for processing once resources in the compute environment are reserved for the job.
10. The method of claim 9, wherein if the party submitting the request for resources has not submitted a job for processing after a predetermined amount of time, then dynamically modifying the request for resources further comprises canceling the request for resources.

11. The method of claim 10, wherein a job comprises one of a reservation, an object that monitors policy, an object that monitors credentials, an object that monitors node states and an object that monitors the compute environment.
12. The method of claim 11, wherein based on the monitored events in the compute environment, modifying the compute environment further comprises dynamically modifying the compute environment to satisfy the request for resources.
13. The method of claim 12, wherein dynamically modifying the compute environment further comprises at least one of: modifying at least one node, modifying at least one operating system, installing end user applications, dynamically partitioning node resources and adjusting network configuration.
14. The method of claim 1, wherein the request for resources is a request for a reservation of resources in the compute environment.
15. The method of claim 14, wherein monitoring events after receiving the request for a reservation further comprises monitoring compute resources associated with the reservation.
16. The method of claim 15, further comprising dynamically modifying the compute environment to more adequately process jobs submitted within the reservation.
17. The method of claim 1, wherein modifying the request for resources comprises migrating a reservation to be associated with new resources.
18. The method of claim 17, wherein migrating the reservation is one of a migration in space and a migration in time to the new resources.
19. The method of claim 17, wherein the new resources better meet needs associated with the request for resources.
20. The method of claim 18, wherein the migration in time seeks to create a reservation at the earliest time possible.
21. The method of claim 18, wherein the migration in time seeks to create a reservation based on availability of resources in the compute environment.

PCT/US05/08298

22. The method of claim 18, wherein the migration in space comprises migrating the reservation to resources that will provide better performance of the compute environment for the request for resources.
23. The method of claim 18, wherein the migration in space comprises migrating the reservation to resources according to a failure or projected failure of resources.
24. The method of claim 1, wherein monitoring events after receiving the request for resources further comprises monitoring a job submitted within a reservation based on the request.
25. The method of claim 24, wherein if the job submitted within the reservation will extend beyond the reservation, the method further comprises canceling the job.
26. The method of claim 25, wherein prior to canceling the job, the method further comprises presenting to the entity that submitted the job an option of extending the reservation to accommodate the job.
27. The method of claim 26, wherein the option of extending the reservation to accommodate the job is subject to pre-established policies.
28. The method of claim 27, further comprising presenting to the entity, with the option of extending the reservation, a pricing option to extend the reservation.
29. The method of claim 1, wherein the request for resources in a compute environment comprises a reservation of resources for a window of time in which at least one user can submit personal reservations.
30. The method of claim 29, wherein personal reservations are one of a non-administrator reservation and an administrator reservation.
31. The method of claim 29, wherein the reservation of compute resources for a window of time is a request for cluster resources for a periodic window of time.
32. The method of claim 31, wherein the periodic window of time may be daily, weekly, monthly, quarterly or yearly.
33. The method of claim 29, further comprising:
receiving a personal reservation for the use of compute resources within the window of time; and

providing access to the reserved compute resources for the personal reservation to process jobs.

34. The method of claim 33, wherein if a received consumption job associated with the personal reservation will exceed the window of time for the reservation of compute resources, then the method comprises canceling and locking out the personal reservation from access to the compute resources.

35. The method of claim 33, wherein if a received consumption job associated with the personal reservation will exceed the window of time, then the method comprises never starting the consumption job.

36. The method of claim 34, further comprising, before canceling and locking out the personal reservation, the step of:

presenting to a user who submitted the personal reservation an option of allowing the jobs running within the personal reservation to complete although it is beyond the window of time for their reservation of compute resources.

37. The method of claim 34, further comprising, if the job submitted under a personal reservation would exceed the personal reservation, extending the personal reservation to meet the needs of the job.

38. A method of managing resources within a compute environment, the method comprising: receiving a request for resources in the compute environment; reserving resources in the compute environment according to the request; and charging the requestor based on the reservation of resources.

39. The method of claim 38, wherein charging the requestor further comprises charging a specific rate for the reserved resources whether the reserved resources are used or not.

40. The method of claim 38, wherein charging the requestor further comprises charging a first rate for reserved resources that are used and a second rate for reserved resources that are not used.

41. The method of claim 40, wherein a used resources is consumed by a job run within the reservation.

42. The method of claim 1, further comprising:

creating a reservation for resources within the compute environment based on the request for resource; and

dynamically customizing resources within the reservation to meet workload submitted within the reservation.

43. The method of claim 42, wherein the reservation is associated with one of an individual or a group.

44. A computer-readable medium storing instructions for controlling a computing device to dynamically manage resources within a compute environment, the instructions comprising:

receiving a request for resources in the compute environment;
monitoring events after receiving the request for resources; and
based on the monitored events, dynamically modifying at least one of the request for resources and the compute environment.

45. A system for dynamically managing resources within a compute environment, the system comprising:

means for receiving a request for resources in the compute environment;
means for monitoring events after receiving the request for resources; and
based on the monitored events, means for dynamically modifying at least one of the request for resources and the compute environment.

46. A system for dynamically managing resources within a compute environment, the system comprising:

a module configured to receive a request for resources in the compute environment;
a module configured to monitor events after receiving the request for resources; and
a module configured to dynamically modify at least one of the request for resources and the compute environment based on the monitored events.

47. A compute environment comprising a plurality of computing devices, the compute environment having resources which are dynamically managed according to a method comprising:

receiving a request for resources in the compute;
monitoring events after receiving the request for resources; and
based on the monitored events, dynamically modifying at least one of the request for resources and the compute environment.